R E P O R T R E S U M E S

ED 013 591 FL 000 533

LET'S LOOK AT RESEARCH.

BY- CHASTAIN, KENNETH
AMERICAN ASSN. OF TEACHERS OF SPANISH AND PORT.

PUB DATE SEP 67

EDRS PRICE MF-$0.25 HC-$0.32 8P.

DESCRIPTORS- *DISCRIMINANT ANALYSIS, *LANGUAGE RESEARCH, *RESEARCH DESIGN, *RESEARCH NEEDS, *STATISTICAL ANALYSIS, EDUCATIONAL RESEARCHERS, RESEARCH CRITERIA, EXPERIMENTS, HYPOTHESIS TESTING, RESEARCH METHODOLOGY, STATISTICAL DATA,

GREATER UNDERSTANDING OF INFERENTIAL STATISTICAL METHODS AND EXPERIMENTAL DESIGN WOULD ENABLE LANGUAGE TEACHERS TO INTERPRET OBJECTIVELY AVAILABLE RESEARCH REPORTS AND HOPEFULLY WOULD ENCOURAGE MORE EXPERIMENTATION. TO UNDERSTAND THE EXPERIMENTAL PROCESS, EDUCATORS MUST REALIZE THAT ANY EXPERIMENT IS AS SUCCESSFUL AS THE EXPERIMENTER IS IN DESIGNING A STUDY ISOMORPHIC TO THE UNDERLYING MATHEMATICAL ASSUMPTIONS AND IN APPLYING STATISTICAL ANALYSIS APPROPRIATE TO THESE ASSUMPTIONS. A RANDOM SAMPLING IS A LOGICAL STEP IN ARRIVING AT THE PROBABILITY VALUES USED IN TESTING A HYPOTHESIS BECAUSE ITS CONSTANT AND INDEPENDENT PROBABILITY FACTORS CAN BE COPED WITH IN SIMPLE MATHEMATICAL TERMS. WITHOUT AT LEAST A CURSORY ACQUAINTANCE WITH RESEARCH REPORTING, EXPERIMENTAL DESIGN, AND STATISTICAL ANALYSIS, EDUCATORS MIGHT EASILY ACCEPT THE EXPERIMENTER'S CONCLUSIONS WITHOUT ANALYZING THE CONTENT THAT LED TO THE CONCLUSIONS. NEVERTHELESS, INTELLIGENT APPRAISAL OF THE LIMITED RESEARCH AVAILABLE REQUIRES A THOROUGH EVALUATION OF ALL AREAS OF INTERNAL AND EXTERNAL VALIDITY. THIS ARTICLE APPEARED IN "HISPANIA," VOLUME 50, NUMBER 3, SEPTEMBER 1967, PAGES 496-500. (AB)

ED013591

A JOURNAL DEVOTED TO THE INTERESTS OF THE TEACHING OF SPANISH AND PORTUGUESE

HISPANIA

FOUNDED 1917

Published by THE AMERICAN ASSOCIATION OF TEACHERS OF SPANISH AND PORTUGUESE

FL 000 533

VOLUME L

September 1967

NUMBER 3

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Vol. L, Number 3

HISPANIA

September 1967

Pages 415-628

LET'S LOOK AT RESEARCH

KENNETH CHASTAIN
Purdue University

Tharp and McDonald in compiling a bibliography of research in foreign language for the *Review of Educational Research* in 1938 included only those articles dealing with research studies. Their findings were quite limited. In the period from December 1933 to 1937 only thirteen articles qualified as reports of true re-

search studies.[1] By 1943 the bibliography had grown to sixty-six references.[2] However, most were opinion articles rather than reports of experimental studies. The criterion of including only research studies had evidently been abandoned. Such seems to reflect to a great extent subsequent bibliographies in the field of modern foreign languages. The majority of the articles appearing in the journals deal principally with personal opinions and preferences.

Certainly the small number of true experiments in the field of modern foreign languages is not to be desired. A wider acceptance and understanding of experimental research and design should be promoted by all those connected with lang study. An increased understanding of inferential statistical methods and experimental design would hopefully lead to an increasing number of experiments among language teachers and at the same time allow them to study much more objectively the research articles now being published. Stanley and Campbell writing in the *Handbook of Research on Teaching* state that they are "gradually coming to the view that experimentation within schools must be conducted by regular staff of the schools concerned, whenever possible, especially when findings are to be generalized to other classroom situations."[3] Obviously some knowledge of research is necessary to conduct experiments, but perhaps not so obviously, and of even greater importance, is the need for understanding by all in order to study the reports of experiments being conducted.

To gain an understanding of the experimental process (without going into the differences between parametric and non-parametric statistics) one must understand that the basis for statistical analysis of any experiment is the underlying mathematical principles. The research project is designed in such a way that the experimental situation corresponds to its mathematical considerations. The experiment then is successful to the point that the experimenter is successful in designing a study isomorphic to the underlying mathematical assumptions and applying the statistical analysis appropriate to these assumptions. In discussing the logic of hypothesis testing Hays describes the experimental process as follows: "from the hypothetical population distribution one obtains a theoretical sampling distribution. Then the obtained results are compared with the sampling distribution probabilities. If the probability of samples such as the one obtained is high, the hypothesis is regarded as tenable. On the other hand, if the probability of such a sample (or one in more extreme disagreement with what is expected) is quite small, then doubt is cast on the hypothesis."[4] The importance of obtaining a random sampling in hypothesis testing is explained by Winer in the the following paragraph: "If a sample is drawn in such a way that (1) all elements in the population have an equal and constant chance of being drawn on all draws and (2) all possible samples have an equal (or a fixed and determinable) chance of being drawn, the resulting sample is a *random* sample from the specified population. By no means should a random sample be considered a haphazard, unplanned sample . . . Random samples have properties which are particularly important in statistical work. This importance stems from the fact that random sampling ensures constant and independent probabilities; the latter are relatively simple to handle mathematically."[5] In other words, random samples are necessary in order to arrive at the probability values used in hypothesis testing.

Keeping the importance of the mathematical assumptions in mind, we can now move toward an examination of the experimental situation itself. Stanley and Campbell have listed the following criteria for evaluating experimental studies:

I. *Internal validity*—This helps us to answer questions as to whether or not the true objectives of the study are really being measured. Were the obtained results due to the effects of the treatments applied in the experiment?

   A. *History*—Without a knowledge of what happens during the experiment we can not be sure whether the results are due to our treatment or to some unrelated variable or variables.

B. *Maturation*—The obtained results might have occurred in the individual's normal developmental process even if no experiment had taken place, i. e., if no treatment had been applied.

C. *Testing*—It is most important to consider the effect of previous testing on a set of second test scores.

D. *Instrumentation*—Changes in the criteria measures such as tests or observers may give different results.

E. *Statistical regression*—No experiment is adequate unless a random sample is obtained. Groups which have been selected from either high scores or low scores will tend to regress toward the average.

F. *Selection*—The experimenter should be careful to avoid bias in obtaining his samples for the comparison groups.

G. *Experimental mortality*—The treatments may effect a differential loss among members of the comparison groups.

H. *Selection-maturation interaction*—An interaction between the selection of the groups and the maturation of group members might affect criteria scores.

II. *External validity*—As well as an examination of the study itself we need to test its applicability. Is this study and group representative of other groups and treatment variables? Would the same results be obtained in other similar groups in the hypothesized population, for example, first year language classes?

A. *Reactive or interaction effect of testing*—Pretests may "increase or decrease respondents sensitivity to treatment and make results unrepresentative for the unpretested universe." Would groups not given the pretests obtain the same results?

B. *Interaction effects of selection biases and the experimental variable*—As a result of selection bias obtained scores may be radically different from what might be obtained from a true random sample of the population or group in which the experimenter is interested.

C. *Reactive effects of experimental arrangements*—The students or groups involved may react to the experimental situation itself. The obtained results would then be inapplicable to groups not in the experiment.

D. *Multiple-treatment interference*—In an experiment with repeated treatments prior treatments may affect results.[6]

One of the most likely dangers is that educators not familiar with research reporting, experimental design, and statistical analyses will merely turn to the last pages of the article and accept the author's conclusions without examining the content which led to those conclusions. Without at least a cursory understanding of the criteria listed above an educator leaves himself open to misinterpretation of the relatively scant number of research studies which have been reported. For example, such a comment as that by Prof. Stack in the *Modern Language Journal* of April, 1964, "Further negative predisposition may be reflected by the 'null hypothesis' to be tested . . . ,"[7] must either be credited to an uninformed person or to one guilty of the same bias of which he had been criticizing another writer. Granting that there is a support position in statistics as well as a rejection position (The traditional position for several reasons has been that of setting up a null or negative hypothesis which is then rejected.), one must conclude that there is no basis in fact for such a statement.

Let's examine a recent, well-known study in the teaching of modern foreign languages, *A Psycholinguistic Experiment in Foreign-Language Teaching* by Scherer and Wertheimer. (The reader should note here that the purpose of the following is to apply the criteria previously given to indicate aspects of any study which may be questionable. The scope of this paper involves merely a sample application of criteria measures and does not allow for a complete evaluation.)

First and foremost, it is questionable whether the mathematical assumptions upon which all statistical analyses rest were met. Before the semester began it was announced that the University of Colorado had been granted government funds to test a new method of teaching modern foreign languages. This announcement together with the connotations accompanying it led to the following statements: "as soon as the students learned that some sections were being taught by an audio-lingual method, many of those in the control group wanted to change to the experimental group." and "The spring-semester registration became somewhat confused because many students tried to register for experimental sections after having been in control sections during the fall semester."[8] Given the above

situation it is doubtful that a random sample upon which both parametric and non-parametric statistics depend could have been obtained.

Now let's turn to an examination of the treatment conditions themselves. The audio-lingual materials were prepared to correspond with the text being used by the traditional class. This attempt to equate the content raises the question of whether it is possible to teach an active command of specified content in the same amount of time needed to teach a passive command. Supposedly both classes covered the same material during the academic year (p. 28).

We can now turn to a consideration of the confounders of internal validity. During a consideration of history several questions arise. (1) The instructors of the experimental and control groups were allowed to teach their preference. Failure to attempt to control the instructor variable could have influenced the results tremendously (p. 4). (2) "Since the same end-of-semester examinations were to be given to both groups, the laboratory was made available to the students of the control group during the last few days of the semester, and copies of the text for the control group were made available to the experimental students at the same time" (p. 26). (3) "Since testing took place in the various sections from January 20 to January 28, there can be no doubt that some students in the later sections received a little information about the tests from students in earlier sections (p. 28). Also, the instrumentation could have affected the internal validity. The authors admitted that it would have been better if the MLA achievement tests had been available (p. 27).

This study is even more suspect with regard to external validity. It is doubtful that the results, due to unusual experimental arrangements not normal to introductory MFL classes can be applied to other situations involving first year language students. In the experiment the students assembled in the evening to take a series of pretests (p. 25). Both the hour of the tests and the quantity were not normal and would have been likely to create a reactive effect on the part of the students. Owing to the student predisposition toward the experimental classes, mentioned previously, one can assume that there were interaction effects of selection biases and the experimental variable and that there were reactive effects to experimental arrangements.

In addition to the above weaknesses one must include an additional one which the authors mention. Tests of significance, t-tests, on the difference between means were used to assess group differences, and in the author's words, "Such statistics are not, strictly speaking, fully legitimate for some of the measures used, since means and correlation coefficients imply a particular degree of refinement of measurement such as interval scaling; by no stretch of the imagination can all our scores be thought to meet these criteria" (p. 2).

The purpose of this article is not to single out one particular study for criticism, but to demonstrate the improvements in our studies which remain to be made. It is disturbing to read and to hear the ready acceptance of the results of many studies in the absence of a true evaluation of the study itself. Too often it seems that the summary of results is accepted without a proper examination of the contents. If a comparison with bowling may be used, we are watching the pins fall without considering the truly important aspect of the game, the approach. Consequently, our progress in the teaching of modern foreign languages may be as "helter skelter" as the fallen pins.

NOTES

[1]J. B. Tharp and K. S. McDonald, "Foreign Languages," *Review of Educational Research*, vol. 8 (1938), 34.

[2]J. B. Tharp, "Foreign Language Instruction: General Review," *Review of Educational Research*, vol. 13 (1943), 115.

[3]Julian C. Stanley and Donald T. Campbell, "Experimental and Quasi-Experimental Designs for Research on Teaching," *Handbook of Research on Teaching*, ed. N. L. Gage. (Chicago, 1963), 191.

[4]Wm. L. Hays, *Statistics for Psychologists* (New York, 1963), p. 153.